

# Novelty in science should not come at the cost of reproducibility

Andrew N. Holding<sup>1,2,3</sup>

1 The Alan Turing Institute, London, UK

2 CRUK Cambridge Institute, University of Cambridge, UK

3 Downing College, University of Cambridge, UK

## Correspondence

A. N. Holding, CRUK Cambridge Institute, University of Cambridge, Robinson Way, Cambridge, CB2 0RE

Tel: 01223 769500

E-mail: andrew.holding@cruk.cam.ac.uk

(Received 12 March 2019, revised 3 May 2019, accepted 19 May 2019)

doi:10.1111/febs.14965

## Introduction

“Non-reproducible single occurrences are of no significance to science.”  
Karl Popper

Oestrogen receptor (ER) cycling is a dogma of the Oestrogen receptor-positive (ER+) breast cancer field [1], which stood unchallenged for nearly 20 years, until recent work by our group [2]. Our aim was to build on the original, highly cited research papers in this field by applying next-generation sequencing methods to expand those previous results from a single gene to the rest of the genome. It sounded like the next logical step, but we quickly ran into difficulty: no matter what we tried, we could not reproduce that early work. In the end, we found ourselves publishing a paper that contradicted, rather than expanded, the original results.

It is hard to believe that I was the first person who had failed to reproduce the previous findings. The underlying problem is that the pressures of modern research reward those who get there first and offer little to encourage people to publish negative results or

even to get it right. These pressures, in turn, distort the literature. We, therefore, need to change the culture into one that enables scientists to take the time to get the right answer – even if it takes multiple attempts – and to recognise that there is more to great science than getting there first.

## Standing on the shoulders of giants

For nearly 20 years, it was accepted dogma that, once stimulated, the ER activated its target genes in successive 90-min ‘on–off’ cycles [1]. That is, until our recent publication [2]. The papers that first detailed this process in 2000 [1] and 2003 [3] currently have over 1000 citations each, yet we could not replicate their key findings. It would be surprising if no one else had encountered this replication problem over the past two decades. However, the real underlying problem here is that negative results are often not published.

## Glossary

ER+ Breast Cancer, The most common subtype of breast cancer – making up 70% of all breast cancer cases – that is driven by the oestrogen receptor; Oestrogen receptor (ER) cycling, A dogma of the ER field, stating that the ER binds DNA in cycles of 90 minutes, resulting in tightly regulated waves of gene activation; ChIP-seq, A method to find where proteins bind genomic DNA; often the binding site gives insight into gene regulation; pfChIP-seq, A method devised by the author to provide additional controls for ChIP-seq that makes use of a ‘parallel factor’ as a control to measure the variability of the ChIP-seq immunoprecipitation step; Immunoprecipitation, The use of antibodies to purify a protein from a complex biological mixture. The method can be highly specific but is complex and therefore often results in high levels of variability; RNA-Seq, A method that monitors RNA levels to detect the expression of individual genes at the genome-wide level.

## Abbreviations

ChIP-seq, Chromatin Immunoprecipitation (ChIP) with next generation DNA sequencing; ER+, oestrogen receptor positive; ER, oestrogen receptor; IP, immunoprecipitation; pfChIP-seq, parallel factor ChIP-seq.

The ER, a key driver of 70% of breast cancer tumours, is a nuclear receptor that is usually inactive in the cytoplasm. Upon activation of the ER by the hormone estradiol, the protein dimerises and translocates to the nucleus, where it directly binds to its target genes to activate transcription. What made the ER special, though, was that in the early 2000s, several papers reported that, once activated, the ER binds to its target DNA and then released it in 90-min cycles, resulting in a tightly timed transcriptional process. In this way, genes were proposed to be turned on and off in cycles.

In 2013, I joined the Markowitz lab as a postdoc, with grand plans to follow up on these studies. The idea was to expand on what had been demonstrated for a single gene in these two highly cited papers by combining it with the latest techniques in whole-genome sequencing. The project proposal was peer-reviewed, assessed to be well-designed, and funded by Breast Cancer Now. However, I quickly encountered challenges.

### **Squinting at data**

Initial results from my own experiments seemed to show the occurrence of the same cyclical signal, as previously reported, but this result could only rarely be reproduced. My immediate reaction was that this was most likely due to how I had carried out these experiments. When thousands of scientists had already cited the previous two reports, it was neither easy nor necessarily the right choice to discount the published data. Yet, with each attempt to reproduce the results, my data continued to tell the same story. Months of work was put into repeating those early experiments, and although they provided hints of success, they never delivered what I considered to be a robust result.

The delays had further implications: the computational biologist with whom I had planned to work with became impatient in their wait for data that was not forthcoming. This person found other projects to work on and then moved on to a new position. The delays in reproducing a previous result were not just having a negative impact on my own research but also on the careers and lives others who were involved in the project.

I worked relentlessly to overcome these challenges. In trying to make sense of the data we had, we tested alternative methods to integrate it and to increase the power of our genome-wide experiments, none of which conclusively reproduced the reported cycling. These attempts were not without some success: one of the methods would go on to be published separately [4]. However, none of these ideas definitively confirmed the cycling result we were looking for.

### **From denial to acceptance to problem solved**

Surrounded by data that contradicted what was published, I started to accept that ER cycling was possibly an artefact. When biased by what the literature told us to look for, we could find it, but when I looked at the data as a whole, the effect vanished.

And from my efforts, I identified a key issue: no one routinely controlled their data for the experimental variability of immunoprecipitation (IP). IP is an important laboratory method that uses highly specific antibodies to selectively bind to target molecules to enable their purification. However, while antibodies are very specific, the complexity of the steps in IP protocols leads to various sources of variability in the final result. The lack of control in each reported ER ChIP experiment meant that, despite the huge volume of work published, you could never conclusively know if the reported changes in ER binding were biologically relevant or occurred as a result of technical variability in the experiment.

A breakthrough came from a collaboration with Michael Guertin at the University of Virginia. Michael had previously worked on ER binding, and he shared my concerns about the lack of rigorous controls in the standard Chromatin Immunoprecipitation (ChIP) with next-generation DNA sequencing (ChIP-seq) protocol, a method that uses IP combined with next-generation DNA sequencing to identify the genomic locations at which specific proteins bind DNA. We worked together to develop a robust method with the controls needed to solve the key variability issue of ChIP. In just over a year, we had jointly published Parallel-Factor ChIP-seq (pfChIP-seq) as a method that would allow us to finally separate the signal from the experimental variation [5].

### **Self-correcting, but at whose expense?**

One of the ideals of science is that it is self-correcting. In public engagement talks, I often use the discovery of a cure for scurvy as an example of how the acquisition of knowledge is not always a simple journey.

Many an aspirational scientist is aware of James Lind's famous experiment in the 18th century, which led to the discovery of citrus fruit as a cure for scurvy [6]. However, by the early 20th century, the leading scientific theory was that scurvy was caused by the bacterial contamination of food. As a result, the measures taken to prevent scurvy were changed and became ineffective [7]. Eventually, research would re-

establish the knowledge of a cure, but it took decades, and over that time, many lives were lost to a relatively curable disease.

What this and many tales like it tell us is that the scientific narrative is rarely linear. We will not always come to the correct conclusion on the first, second or third attempt.

The often-convoluted path that science takes is not a problem with the method *per se*. However, we need to insure against the impact of irreproducible research on individuals (the ‘human’ impact), particularly when jobs and livelihoods depend on the ability to replicate what’s previously reported. I was fortunate that, due to being in a secure position, I could take my time and do the research thoroughly with the support of my peers.

Given the impact that irreproducible research can have on researchers’ careers, I believe that we need a more compassionate career path that takes into account that scientific research does not happen without personal cost: young scientists might need to financially support their families or be subjected to visas that are conditional on their studentship. Running out of time can mean that a student is not awarded a PhD nor gets to publish that crucial paper.

These challenges are combined with the pressure on early-career researchers to deliver high-impact results. The outcome is an environment that pushes people to get across the line as quickly as possible, while the incentives to challenge or to reproduce previous studies are minimal.

### The fallacy of sunk costs

We all have a bias to overvalue what we have invested in, sometimes heavily, because we are tainted by the emotional investments we accumulate. This is referred to as the ‘the fallacy of sunk costs’: the more we invest in something, the harder it becomes to abandon it. And this bias still holds even when we know that an alternative can deliver more for less by way of future investment. After the publication of the pfChIP-seq method [5], I was especially aware of the risk of following my initial project further. Was I just a gambler chasing my losses?

The time and money invested in that project had been significant, and the process had taken its toll on some of those working on it. I also had several other projects in the pipeline that were going well and I had no idea how much push back I would get at peer review if I were to try and get my results published.

I was fortunate. I knew my funding was secure, mitigating some of those risks. But even though I had

absolute confidence in the methods I had developed, I knew that we had to get it right first time. Where ChIP-seq experiments often have two replicates, I included six. Where most ChIP-seq experiments have two conditions, I had 10 time points. These choices all came with added cost, building on the ever-increasing pressure to get a publishable result.

### Confirmation Bias

The experience and time that led up to our final experiment paid off. Data analysis enabled us to clearly see that our controls were stable. The challenge in reproducing ER cycling genome-wide was not due to my experimental error; it simply was not there in the data.

It would have been much easier to pick the samples or the data that gave the effect I was looking for, but in the end, my doubts were justified. The cycling effect my brain kept seeing – only hinted at in my data – was just confirmation bias.

The reviewers’ comments are available with the resulting publication [2] and are well worth reading. The response to the original submission shows that one reviewer had experienced the same challenges as I had had when trying to replicate the reported results. Our eyes are drawn to seeing patterns, even when they are not there. When followed up, none of it was significant.

In all likelihood, my result will not be the final one in this story. Within the publication, I stated ‘we would welcome further replication of this study’, and I still hold by that statement. It is entirely possible that what I have found in a cell line might be very different to what occurs in a patient. There is always the possibility that a future technology will lead to a study that supersedes my own results. This is a humbling reality that we must accept as scientists but not one that should limit us.

### Reproducibility as a Philosophy

Kirstie Whitaker, a Research Fellow at the Alan Turing Institute (London, UK), spoke at the Turing Institute Health Programme Conference in Manchester, in March 2019 [8]. She discussed the current barriers to reproducible research. Her belief is that the key problem is not fraud, but instead the pressure that the culture of science places on the individual. The risk of embarrassing, high-profile retractions also prevents data from being published that could correct the published literature. Such avoidance also enables researchers to plead ignorance if they have made a genuine mistake. Meanwhile, the perceived time and

cost of working reproducibly and its lack of requirement for career progression select against, rather than promote, a culture that supports scientists to work in this way.

There is a growing acceptance of the points Kirstie raises as key challenges facing research excellence. The Royal Society highlighted similar issues as part of their 'Changing Expectations' programme, citing the hyper-competitive research culture and the current narrow definitions of success as primary drivers for the challenges I describe [9]. Likewise, the 2014 review by the Nuffield Council of Bioethics [10] provided considerable evidence for these claims, concluding that the current culture 'create[s] incentives for poor quality research practices, less collaboration and headline chasing'.

Personal responsibility for reproducible science is growing, and strong arguments have successfully been made that it is in one's own interest to strive for reproducible data [11]. However, we need more than this. We need to make reproducing the work of others a key part of a scientist's career, and not just something done in an *ad hoc* fashion by the privileged few who can afford it.

The pressure to be novel and first is systemic and troubling. Despite attempts to correct the distortions in the evaluation of scientific research [12], the careers of scientists are still often assessed and rewarded on the basis on where they publish rather than what they publish. Or as Brian Nosek, Executive Director of the Center for Open Science (Virginia, USA) concisely states: 'Incentives for individual success are focused on getting it published, not on getting it right' [13]. A recent study by Stephanie Hicks [14] showed that many of the early single-cell RNA-sequencing experiments mistake variability for novel results, with some studies showing 100% of the results derived from confounding factors. Without releasing the pressure on individuals, the way in which researchers are currently assessed and incentivised will continue to distort the published literature. It would be preferable to have the time to do it right, while simultaneously allowing scientists to be human and make mistakes, instead of focusing on novelty, being first and publishing in highly selective journals.

## Where next?

Programmes like the reproducibility project for cancer biology provide part of the answer [15]. By preregistering the experiments, the outcomes – positive or negative – are always reported. Scientists will be credited for their work either way, thereby combating the

inherent publication bias within a system that might promote the one time that an experiment gives a particular result, rather than all of the unpublished evidence that it does not. While currently limited in numbers, in the future, these studies should form an essential part of every scientist's career.

We need to challenge the culture of 'first past the post' and the pre-eminence given to high-impact journal publications. Instead, we need to develop a research culture in which research outcomes are only seen as being groundbreaking once they have been independently validated. In such a culture, those who replicate research findings are seen as being an important part of the research process – as important as those who got there first. In my case, so few had the time and resources to challenge the established fact that they simply published the next novel result without ensuring the underlying model was right.

Reproducible science also means greater impact. In my example, the ER is of interest as a target for several key therapeutics used to treat breast cancer [16]. Understanding the mechanism by which it functions plays a crucial part in our strategies for future interventions. However, those future outcomes are only as good as the evidence that we base them on.

Whatever model we come up with, it does not need to limit the environment that has already led to promising young scientists producing high-impact science. We should continue to value their significant contributions. We are already seeing back-to-back submissions to journals, when by chance, two groups have found the same answer from two different directions. In terms of the science, this is a perfect solution, reproducible and independent. We need to shift to a scientific culture where this happens more often. And, rather than fearing 'being scooped' – when another scientist through independent research publishes key findings before you – we should see it as something to embrace and promote.

What I planned to do seemed like a straightforward next step for the field. It would have been reasonable to assume that the labs who reported cycling had all the resources to undertake the work themselves; in hindsight, one could say that the lack of follow-up publications should have been a warning sign. Discussions I have since had at conferences confirm that this was exactly the case, with groups struggling to reproduce the results genome-wide. Nonetheless, none of that insight or those conversations were possible without the journey I took. In light of that, the only advice I have for my past self is: 'you have to follow the data, even if you don't like where it goes'.

## Acknowledgements

This work was supported by the Alan Turing Institute under the EPSRC grant EP/N510129/129/1 as a Turing Fellowship to ANH.

## Conflict of interest

The authors declare no conflict of interest.

## References

- 1 Shang Y, Hu X, DiRenzo J, Lazar MA & Brown M (2000) Cofactor dynamics and sufficiency in estrogen receptor-regulated transcription. *Cell* **103**, 843–852.
- 2 Holding AN, Cullen AE & Markowitz F (2018) Genome-wide Estrogen Receptor- $\alpha$  activation is sustained, not cyclical. *eLife* **7**, e40854.
- 3 Métivier R, Penot G, Hübner MR, Reid G, Brand H, Koš M & Gannon F (2003) Estrogen receptor- $\alpha$  directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* **115**, 751–763.
- 4 Holding AN, Giorgi FM, Donnelly A, Cullen AE, Nagarajan S, Selth LA & Markowitz F (2019) VULCAN integrates ChIP-seq with patient-derived co-expression networks to identify GRHL2 as a key co-regulator of ER $\alpha$  at enhancers in breast cancer. *Genome Biol* **20**, 91.
- 5 Guertin MJ, Cullen AE, Markowitz F & Holding AN (2018) Parallel factor ChIP provides essential internal control for quantitative differential ChIP-seq. *Nucleic Acids Res* **46**, e75–e75.
- 6 Lind J (1757) A Treatise on the Scurvy: In Three Parts, Containing an Inquiry into the Nature, Causes, and Cure, of that Disease. A. Millar, London.
- 7 Ceglowski M (2001) Scott And Scurvy. [https://idleword.com/2010/03/scott\\_and\\_scurvy.htm](https://idleword.com/2010/03/scott_and_scurvy.htm)
- 8 Whitaker K (2019) The Turing Way: Reproducible, Ethical and Collaborative Data Science. figshare. Presentation. <https://doi.org/10.6084/m9.figshare.7819442.v1>
- 9 The Royal Society (2018) What is Research culture? *YouTube*, <https://www.youtube.com/watch?v=mZ3bdTmjPKg&feature=youtu.be>
- 10 Nuffield Council on Bioethics (2015) The culture of Scientific Research in the UK, *Web Report*, <http://nuffieldbioethics.org/project/research-culture>
- 11 Markowitz F (2015) Five selfish reasons to work reproducibly. *Genome Biol* **16**, 274.
- 12 Alberts B (2013) Impact factor distortions. *Science* **340**, 6134, pp. 787
- 13 Nosek BA, Spies JR & Motyl M (2012) Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Pers Psychol Sci* **7**, 615–631.
- 14 Hicks SC, Townes FW, Teng M & Irizarry RA (2017) Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostatistics* **19**, 562–578.
- 15 Errington TM, Iorns E, Gunn W, Tan FE, Lomax J & Nosek BA (2014) Science forum: an open investigation of the reproducibility of cancer biology research. *Elife* **3**, e04333.
- 16 Jordan VC (2003) Tamoxifen: a most unlikely pioneering medicine. *Nat Rev Drug Discov* **2**, 205–213.